



-1-

501  
IMPROVED EXPRESSION OF GENES IN PLANTS

Field of the Invention

5 The present invention relates to the general field of genetic engineering and is directed, in particular, to improvements in the coding sequence for foreign genes to be expressed in the cells of higher plants.

Background of the Invention

10 It is now possible reliably and repetitively to insert foreign genes into the germ line cells of higher plants, at least for certain species. A variety of techniques exist, notably Agrobacterium-mediated plant transformation and particle-mediated plant transformation, by which foreign genes can be introduced into the germ line plants in such a fashion  
15 that progeny of the plants will bear the gene of interest which is inserted. Accordingly, one area of research directed toward the creation of improved transgenic plants of potential commercial interest is in the insertion into plants of useful genes obtained  
20 from other species or classes of organisms so that the benefits of the gene product can be imbued into certain lines of higher plants. Examples of gene products in which effort has been directed toward their expression

in plants cells include various toxins for control of insects, genes coding for various kinds of viral or other pathogen disease resistance, and genes coding for resistances to specific herbicides or antibiotics. In many of these cases the gene which is desired to be expressed in the plant cell comes from a procaryotic or viral organism. Some foreign genes may be from other species of plant or from other plants of the same species. When heterologous genes from these sources are inserted into plants, using promoters and expression cassettes which have been found operable and effective to express genes in plant cells, the results have been found to be sometimes uneven. There are apparent differences in either the transcription or translation levels of given coding sequences in plant tissues, even if the coding sequences are under the control of identical transcriptional promoters and terminators.

An example of this phenomenon has been found to occur with the gene for the delta-endotoxin crystal protein gene from the soil dwelling microorganism Bacillus thuringiensis (hereinafter referred to as the B.t. gene). A number of B.t. genes coding for homologous proteins have been cloned and sequenced by a variety of investigators throughout the world. Several of genetic constructs including one of the B.t. genes have been used to create chimeric plant expression gene constructions which are then transferred into the cells of plants. The various B.t. genes have been found to have significant differences in the DNA coding regions of the genes, although there is relatively high homology in the proteins for which they code. Nevertheless, the B.t. genes have characteristically been found to express relatively poorly in plant cells as compared to most other gene products which have been introduced into the cells of higher plants. The phenomenon of poor or low expression appears to have

been experienced in all examples to date resulting from the introduction of native coding sequences for B.t. genes into plants, even though the expression cassettes and promoters and transcription terminators varied from experiment to experiment. One possible explanation for the observed phenomenon might be some feature of the native bacterial coding sequence itself.

As is known to all of ordinary skill in molecular biology, the genetic code of three nucleotide units, or codons, specifying particular amino acids, is degenerate. While a single amino acid is specified by each three nucleotide codon which makes up the genetic code found in DNA or RNA, because there are less amino acids possible than there are codon arrangements possible, most amino acids are specified by more than one codon sequence. For example, the amino acids serine, arginine, and leucine are all specified by any of six possible codons. It is thus possible to have nucleotide coding sequences for proteins which can differ significantly in their nucleotide sequence while specifying an identical amino acid sequence for the resultant protein.

#### Summary of the Invention

The present invention is summarized as a method for constructing chimeric coding sequences for expression in plant cells in which the native coding sequence for a foreign gene to be expressed in plant cells is modified by substituting for the codons in the foreign coding region codons which are preferentially expressed in plants. The codons preferred for expression in plants are determined by analysis of the codon usage pattern of plant genes which are natively efficiently expressed in native plant tissues.

The present invention is further summarized in that a plant is engineered with a chimeric gene construct including a protein coding region

690

constructed, and least in part, by oligonucleotide synthesis wherein the oligonucleotides are selected on the basis of preferred codon usage as determined by the usage of codons in genes which express well natively in plants.

It is an object of the present invention to enable the efficient construction of plant genes so as to obtain high steady-state levels of transcription and expression.

It is another object of the present invention to provide a B.t. gene construction which provides for high steady-state level of transcription and expression of the B.t. delta endotoxin protein in plant cells.

Other objects, advantages, and features of the present invention will become apparent from the following specification when taken in conjunction with the accompanying drawings.

#### Brief Description of the Drawings

Fig. 1 is a table of preferred codon usage for use within the practice of the present invention as described further below.

Fig. 2 is a comparison of the coding regions of pAMVBTS and pAMVBT4.

Fig. 3 illustrates the sequence and assembly of oligonucleotides KB72 and KB73.

Fig. 4 illustrates the sequence and assembly of oligonucleotides KB74 and KB75.

Fig. 5 illustrates the sequence and assembly of oligonucleotides KB76 and KB77.

Fig. 6 illustrates the assembly of the oligonucleotides and their insertion into pAMVBTS.

#### Detailed Description of the Present Invention

The principle of the present invention is based on an insight derived from scientific investigation into the problem of expressing significant levels of the

B.t. gene in plant cells. As already mentioned above, previous reports of the creation of chimeric expression constructions including the B.t. gene and the introduction of those constructions into the genome of plants have given rise to relatively low levels of expression and low levels of measurable mRNA. It can be demonstrated that a B.t. expression construction using a plant promoter known to work well with other genes, such as the cauliflower mosaic virus 35S (CaMV 35S) generates a lower steady-state level of mRNA in plant cells than other genes inserted behind the same promoter. Since the problem appeared to be organic to the native B.t. coding sequence itself, the nature of that coding sequence itself was investigated in detail. Such analysis revealed one feature in particular that seemed to be a relatively unique feature of all reported B.t. genes. All of the reported native B.t. genes seem to have a high proportion of A and T nucleotide basis in their coding sequence, relative to other bacterial coding sequences that had been found to be more easily expressed in plants. The reason for this is obscure. Nevertheless, it seems that different coding sequences coding for identical proteins could have differing characteristics of mRNA stability or of interaction with the translational machinery of a given type of cell. For example, the chemical binding energy and secondary structure of mRNAs can be different depending on the relative proportions of the nucleotide pairs. It is also quite possible that the nucleotide content of a given mRNA may affect the strength of the interaction of that mRNA with ribosome.

Regardless of which of these, or if other, theories are appropriately correct to explain the difference in nucleotide content of the B.t. gene from other genes which express well in plant cells, one could logically assume that the plant transcriptional and translational systems which had evolved over time

within the cells of plants themselves would have evolved to have some optimal or increased efficiency for those genes important to the plant system itself. Thus it becomes unnecessary to understand the exact system by which an mRNA of a certain nucleotide content might be preferred over an mRNA with a different nucleotide content, if the phenomenon can be used to advantage simply by examining the coding regions known to express well in plants to determine the nucleotide and codon usage characteristics of those molecules.

To determine those codons which are therefore "preferred" for usage in plant cells, or those which are preferentially expressed in plant cells, it was determined that a logical place for inquiry would be plant genes themselves, to the extent they are known. Certain public sequence data base services (for example GenBank) contained within them many sequences for plant genes which have been sequenced and had their sequence published. It is therefore possible to examine those published sequences to determine within those plant genes which codons are preferred compared to others which are not preferred. In order to accomplish this objective, the GenBank and EMBL public sequence data bases were utilized. In order to correct for possible bias due to the over representation of certain kinds of genes, within the limited number of plant gene sequences which are contained in present data bases, a number of limiting assumptions were made in the compilation. A tailored list of genes was created intended to avoid placing over emphasis on the families of genes which have been most studied. Therefore, for example, only a limited number of storage protein genes were included within the information base on codon usage. A representative storage protein gene was selected from each of maize, soybean and other important crops, and the remaining storage protein genes were considered not to be distinct from these

representative sequences. Other gene types which were also over represented in the publicly available data bases, such as heat shock genes, were similarly selected from. The information was further edited to  
5 include only complete coding sequence information where available. Information was pooled into one common information base, regardless of plant species from which the gene sequence was derived. Data was not species specific only because there are not sufficient  
10 numbers of reported gene sequences from any one given plant species of interest to be sufficiently statistically useful in and of itself. Data from genes that express in different tissues or different periods of development, but are similar, were also pooled on  
15 the theory that there are not enough examples in the kinds of genes available to provide a significant consensus sequence.

As research in the molecular biology of plant genes continues, the knowledge base of published plant  
20 gene sequences may expand to the point where more specificity in determining preference of codon usage may be possible. For example, it may develop that certain plant species may have a preference for a given pattern of codon usage over that pattern preferred by  
25 another species. There may also be differences in codon usage among cell or tissue types in the same species. Thus, while the tabulation of plant codon usage developed here is generally useful and probably a good approximation of an optimum pattern of usage for  
30 plants in general, it may be preferred to a given tissue or plant to have a modified table of codon usage more specific to that tissue or plant.

Once the information base of publicly available plant gene sequences was assembled, a codon usage table  
35 for plant genes in general was compiled by an appropriate computer program, which analyzed all of the codons used in all of the plant gene sequences

contained in the information base. The table representing the results of this compilation is contained in Fig. 1 herein. This table shows the frequency of use of the various plant codons contained within the information base generated from the publicly available plant gene sequences. The farthest right number associated with each codon is the percentage that that codon is utilized by the plant gene sequences in the public sequence data base as a proportion of all of the codons which code for the same amino acid. Thus, for amino acids for which there is only one codon, such as methionine and tryptophan, the codon has a usage factor of 1.0 indicating that it is used all the time when that amino acid is specified. As another example, for the amino acid aspartate, the codon GAT is used 45% of the time that the amino acid is specified in the total of all the plant genes in the information base, while the alternative codon for aspartate, GAC, is used at a frequency of 55% of the time of the coding sequences in the data base.

An examination of the usage table contained in Fig. 1 reveals strong biases in codon usage among the plant genes for several amino acids that have degenerate codons for the same amino acid. As an example, for the amino acid lysine, in plant genes 81% of the time where the amino acid is to be specified, the codon AAG is utilized while only 19% of the time that the amino acid is to be specified is the codon AAA utilized. As another example, of the six possible codons which code for the amino acid leucine, four of the codons represent 92% of the total leucine codon usages, while the two codons TTA and CTA are used a total of only 8% of the occurrences of a leucine codon within the coding sequences of all of the plant genes in the information base. Similar biases, which vary in strength, are present for almost all of the amino acids.

74



It was then possible to compare the codon usage for the native B.t. coding sequence with the codon usage frequency of native plant genes. The results were quite striking, in that in most instances where the table of preferred codon usage for plant genes shows a bias toward a particular codon usage, the native coding region for the B.t. gene showed precisely the opposite preference of use. As an example, for leucine, a preferred codon found in the native coding region of the B.t. gene was the codon TTA, which appeared 45% of the time that the amino acid leucine was to be specified by the B.t. gene, while that codon is the least preferred of all of the possible leucine codons in plant genes, representing only 3% of the total codon usage. In the native B.t. coding sequence it was determined that the twenty-six TTA leucine codons represented 4% of the total of the amino acids in the protein which indicated that the native coding region for the B.t. gene is not typical of what is found in a native plant gene. An examination of other chimeric constructions including other bacterial genes which have been found to express well in plants, no similar problems could be uncovered. Most gene products which have been found to express well in plants conformed well to the plant codon usage table, with there seeming to be some correlation between the level of expression and the highest correlation to the codon usage preferred by plants as represented by the codon usage table of Fig. 1.

Using this data it was then possible to construct a synthetic B.t. coding region for a chimeric gene composed principally of codons selected from those codons which are preferentially expressed by plants as determined by the usage pattern of plants illustrated in Fig. 1. Rather than synthesizing the entire coding region of the B.t. gene, it was first decided to synthesize the 5' end of the coding sequence, and to

determine the effect of the codon substitution in that region on the overall expression of the gene product by the plant cells. Therefore, using the table of preferred codon usage as a guide, a nucleotide sequence was designed for the first 138 codons of the B.t. coding region. The codons for each codon set of this synthesized B.t. region were selected to code for the identical amino acids present in the native procaryotic protein, but were selected to be the particular codon that had the highest frequency of use according to the plant gene codon analysis described above. In other words, the chimeric nucleotide coding sequence was specifically constructed to code for the expression of the same amino acid but was made up of codons different from that in the native organism and selected from those codons determined to be preferentially efficiently expressed by native plant genes. These changes were made on a pre-existing B.t. expression plasmid, referred to as pAMVBTS, previously used by the inventors here to express the B.t. gene in plants. Fig. 2 attached hereto shows a sequence comparison of the original coding region for nucleotides 480 through 903 of the pAMVBTS gene aligned with the synthetic coding region specified as described above. Nucleotide homologies between the two sequences are noted. The sequence in pAMVBTS is the sequence natively present in the HD-1-DIPEL subspecies Kurstaki gene of Bacillus thuringiensis. It is a feature of this alignment table that it can be seen that many of the nucleotides in the third position of the codon have been altered. This is to be expected since the third position is the most degenerate position to conserve amino acid code. The most frequent change of actual individual nucleotide is from an A in the third position in the native procaryotic sequence to another nucleotide, usually C or G, in the chimeric synthetic sequence. The overall

effect of the changes was an increase in C and G content and a decrease in A and T content.

Since the synthesis of an oligonucleotide exceeding 400 base pairs in length is rather difficult, the actual synthesis of the synthetic coding region, described below, was constructed by constructing six separate oligonucleotides which were composed of three separate overlapping pairs. The overlapping pairs were hybridized and then extended into complete duplexes by Klenow polymerase. The three sets of oligonucleotides were arranged so that they would be easily annealed end-to-end to create the entire synthetic coding region. The sequence of the particular oligonucleotides is given in the attached drawings so that construction of these same oligonucleotides can be accomplished by those skilled in the art.

The synthetic coding region thus constructed serves as a protein coding region which can be combined with flanking regulatory sequences for creating a chimeric gene for transformation into a plant to create transgenic plants expressing the B.t. protein. Any otherwise suitable regulatory sequences, such as promoters, 5' non-coding sequences and polyadenylation sequences, are effective with this coding region. The chimeric gene may be inserted through any conventional transformation technique into any plants capable of transformation. While the results indicated below have been conducted with the model species tobacco, the use of tobacco is principally as the result of the ease of transformation and regeneration of tobacco plants, thus making it relatively easy to achieve transgenic expression. Results with the native B.t. coding region have indicated that expression cassettes active to express the B.t. coding region in tobacco are similarly active in cotton and in other plants. Since the preferred codon usage table of Fig. 1 was derived by reference to all plants, rather than just tobacco,

there is good reason to believe and expect that the increased efficiency of expression achieved in tobacco through the use of the method and coding region of the present invention will be equally applicable in other plant species, as it is in tobacco, as demonstrated by the results here.

It also becomes obvious to one skilled in the art that the method is used with the particular procaryotic gene described and illustrated in the present invention is equally applicable to other procaryotic or even eukaryotic, genes which happen not to express well in plants. The results of this procedure demonstrate that at least one factor in the relatively low expression level of the procaryotic B.t. protein in plants is due to the actual makeup of the codon usage pattern of the particular procaryotic gene. Other procaryotic or eukaryotic genes which similarly use a large number of codons which are not among those preferentially expressed by plants may also be altered in the similar fashion. Again the actual protein made by the plant can be identical in the amino acid sequence to the protein encoded by the native foreign gene. Only the codons are switched, not the amino acid that is coded. Therefore it is possible to express many foreign proteins effectively and efficiently in plant cells and still to produce a protein identical in amino acid sequence to the native protein while still gaining the efficiencies possible using the transcriptional and translational machinery of plants more effectively.

This method may even be applicable to some plant genes. It can be readily imagined why some plant genes may be advantageously expressed at less than total efficiency, and one mechanism which might be used is inefficiencies in the pattern of codon usage. As an optimal pattern of usage is developed, it may be possible to enhance the level of a native plant gene by similarly changing the pattern of its codon usage and

returning the modified gene to a plant of the same or different species.

As an examination of the following Examples will reveal to one skilled in the art, the substitution of plant preferred codons in a plant expression cassette results in an increased level of efficiency in expression of the engineered protein. In the following example, the coding region of the protein expression cassette was altered by as few as 59 to as many as 138 codons, all at the amino terminal end of the protein or the 5' end of the coding region. Since the results did not seem to vary greatly based on the length of the substituted codons, it is possible that the increased expressional efficiency is due principally to the substitutions at the amino-terminal, or 5', end of the coding sequence, perhaps those in the first 25 codons. One possible explanation for this might be increased efficiency in binding to ribosomes. If true, this would suggest that entire coding regions need not be altered to gain a relatively significant increase in efficiency of expression, merely the amino-terminal end of the coding region, for perhaps about 25 codons. Performing such a codon substitution for the remaining portion of the coding region might still be expected to increase efficiency of expression, although perhaps less dramatically.

The present invention will be understood to be more generalized from a consideration of the following example of the practice of this invention.

#### 30                    Examples

As described above, a chimeric synthetic coding sequence for the first 138 codons of the B.t. gene coding sequence was constructed. This coding sequence was constructed by synthesizing six oligonucleotides which were grouped in three overlapping pairs. Each single stranded oligonucleotide was then hybridized to

its partner which it overlapped. The two joined  
oligonucleotides, now partially double-stranded, were  
extended into complete duplexes through the use of  
Klenow polymerase. The oligonucleotide pairs were  
5 designed to have overlapping 3' ends in each pair to  
form priming sites for the action of the polymerase.  
The ends of each pair were designed to include  
restriction sites for efficient joining of the ends of  
the double-stranded oligonucleotides together into the  
10 B.t. expression plasmid.

The process began with the construction of the six  
oligonucleotides. The complete sequences for all six  
oligonucleotides and their assembly into the three  
double-stranded coding region segments is illustrated  
15 effectively in Figs. 3, 4 and 5. The particular  
oligonucleotides were designated KB72-KB77. As  
illustrated, for example, in Fig. 3, the  
oligonucleotide KB72 was constructed so as to have a  
complementary 21 nucleotides to the end of the  
20 oligonucleotide KB73. The two oligonucleotides were  
then annealed and extended with a Klenow polymerase  
plus four deoxynucleotide triphosphates. The annealed  
double-stranded DNA was then processed through a phenol  
extract to inactivate the Klenow polymerase and was  
25 digested with Nco I and Spe I to reveal the sticky ends  
illustrated in Fig. 3. Similarly as can be seen with  
reference to Figs. 4 and 5, the oligonucleotides KB74  
and KB75 were annealed, extended, and digested to  
result in a fragment having sticky ends resulting from  
30 digestion by the Ban I and Xba I and the  
oligonucleotides KB76 and KB77 were hybridized,  
annealed and digested to result in a fragment having  
sticky ends digested by Xba I and Bsp 1286.

The assembly of the three coding sequence  
35 fragments into the genome of pAMVBTS was constructed in  
three stages resulting in the sequential construction  
of three plasmids, pAMVB2, pAMVB3 and pAMVB4, each

20

one of which had a sequentially greater portion of its coding region substituted by the synthetic sequence. The process began with the plasmid pAMVBTS as illustrated in Fig. 6.

5        Before insertion into the actual expression plasmid, the three blunt ended duplex fragments were first cloned into pUC12 and the synthetic DNA was sequenced to confirm that the synthesis had been correct. The synthetic inserts were freed from pUC12  
10      by preparative digestion of the plasmids with the appropriate restriction enzymes to generate the required sticky ends. The fragments were purified from agarose gels.

15       The plasmid pAMVBTS was digested with Nco I and Spe I and the vector was purified away from the small 178 nucleotide fragment which had been excised from the plasmid. The synthetic fragment containing both KB72 and KB73 was then ligated with the larger portion of the pAMVBTS vector and the E. coli strain MM294 was  
20      transformed to ampicillin resistance. The resulting plasmid pAMVBT2 was identified by minipreps. This plasmid, pAMVBT2 was thus a complete plant expression plasmid containing the 35S promoter from cauliflower mosaic virus, a 5' non-coding region from the alfalfa  
25      mosaic virus, a B.t. coding region coding for the approximately 72 kilodalton Amino-terminal toxin portion of the native Bacillus thuringiensis delta endotoxin protein, but which differed from the native sequence by the substitution of the original native 59  
30      codons with codons preferred by plants, followed by a polyadenylation sequence derived from nopaline synthase.

35       The plasmid pAMVBT2 was then digested with Ban I and partially digested with Xba I and the vector was purified to remove 132 base pair fragment released by these enzymes. The synthetic fragment formed from the oligonucleotides KB74 and KB75 was ligated to this

vector and transformed into E. coli strain MM294 which was transformed to ampicillin resistance. The plasmid pAMVBT3 was identified by miniplasmid screening.

5 Annealing of this insert into the larger portion of pAMVBT2 destroyed the Spe I site used in the construction of pAMVBT2. The amino acid specified by the Spe I recognition site did not conform to the codon usage table as specified by the preferred codon usage table of Fig. 1, but was a convenient site to retain  
10 until the construction of pAMVBT3. The plasmid pAMVBT3 was similar in all respects to pAMVBT2 with the exception that the substitution of codon usage from the native sequence had been extended for another 45 codons as compared to pAMVBT2.

15 To construct pAMVBT4, pAMVBT3 was first digested with Xba I and Cla I. The resulting 3,589 base pair fragment including the amino and carboxyl-termini of the B.t. toxin coding sequence and the rest of the expression cassette was purified away from the two  
20 smaller fragments, of 619 and 375 base pairs, released by the double digestion with these enzymes. The plasmid pAMVBT3 was then digested in a second reaction with Bsp 1286 and Cla I and the small fragment corresponded to the internal region of the B.t. toxin  
25 coding sequence between nucleotides 897 and 1767 with Bsp 1286 and Cla I sticky ends was purified. A ligation reaction was then conducted between the 3589 base pair vector from pAMVBT3 plus the 870 base pair coding region of pAMVBT3 (from the Bsp1286 site to the  
30 Cla I site) and the synthetic duplex of KB76 and KB77. The resulting plasmid was transformed into E. coli strain MM294, which was selected for ampicillin resistance, and the desired plasmid pAMVBT4 was again identified by plasmid minipreps.

35 Each of the plasmids pAMVBT2, pAMVBT3 and pAMVBT4 were individually co-integrated into the carrier plasmid pTV4. The plasmid pTV4 is contained within a



plasmid pTV4AMVBTSH, which is ATCC Accession Number 53636, and can be readily retrieved from this plasmid by digestion with Xho I to completion, phenol extraction and ethanol precipitation after which the resulting plasmids can be religated, transformed into E. coli, and selected for sulfadiazine resistance. The sulfadiazine resistant colonies will contain the plasmid pTV4.

The plasmid pTV4 is a carrier plasmid containing a unique Xho I site bounded in one direction by a synthetic consensus right border sequence similar to the right border of T-DNA from Agrobacterium tumefaciens, and in the other direction, a complete expression cassette for the kanamycin resistance trait as conditioned by the plant expression gene APH-II, and a synthetic consensus left border sequence similar to the left border of Agrobacterium T-DNA. The plasmids pAMVBT2, pAMVBT3 and pAMVBT4 can be digested at their unique Xho I site, which is 5' to the coding region for the B.t. expression cassette, and ligated into copies of pTV4, also digested with Xho I, to result in complete transformation cassette, including the B.t. coding gene for kanamycin resistance, and left and right T-DNA borders suitable for transformation into plants.

These co-integrations were constructed and the three resulting transformation plasmids were conjugated into A. tumefaciens strain EHA101 in a manner similar to that described in Barton, et al., Cell, 32, pp. 1033-1043 (1983). Seeds of tobacco were surface sterilized and germinated on Murasige and Skoog (MS) medium. Aseptically grown immature stems and leaves were then inoculated with overnight cultures of A. tumefaciens harboring the appropriate transformation plasmid. Following 48 to 72 hours of incubation at room temperature on a regeneration medium (MS medium containing 1 micrograms per ml of kinetin), cefotaxime

(at 100 micrograms per ml) and vancomycin (at 250 micrograms per ml) were applied to kill the Agrobacteria, and kanamycin (at 100 micrograms per ml) was applied to select for transformant plant tissues.

5 After approximately six weeks, with media changes performed at two week intervals, shoots appeared. The shoots were excised and placed in rooting medium containing 25 micrograms per ml kanamycin until roots

10 roots were formed, which occurred in 1 to 3 weeks. After roots were formed, the plants were transferred to a commercial soil potting mixture for growth into mature plants. Insect toxicity tests were conducted on leaves of the resulting whole, intact, although small, tobacco plants.

15 Insect eggs of tobacco hornworm (Manduca sexta) were hatched on mature, wild-type tobacco plants. Larvae of the insects were allowed to graze for 1 to 3 days on wild-type plants prior to transfer to test plants. Since mature tobacco plants contain higher

20 levels of secondary metabolites than freshly regenerated plants, the feeding of the larvae on the older plants made the larvae less sensitive to toxins than neonatal larvae. This was done to reduce the sensitivity of the larvae and this distinction proved

25 useful in distinguishing between variations in the toxin produced in the transgenic plants. Tobacco hornworms were placed directly on the leaves of the young wild-type plants and on recombinant plants in number of 2 to 4 larvae per plant per test with up to 6

30 successive tests conducted per plant. Tests were conducted and the plants were graded as to their toxicity to the larvae. The plants were considered to be "killers" if all of the larvae grazing on the leaves of the plants ultimately terminated. The plants were

35 rated relative to each other on the length of time and degree of feeding necessary before the "killer" plants caused death of the hornworms. A rating of "9" was

64  
6

indicative of a strongly resistant plant, where the high level of toxin present caused rapid cessation of feeding and early death. A rating of "5" or less indicated moderate toxicity, in which generally one or  
5 more days of limited feeding occurred before larval death.

Shown in Table I is a summary of the results of the hornworm feeding trials conducted with these three plasmids as compared to the plasmid pTVAMVBTSB which  
10 contains the native coding sequence derived from the native bacteria. The results illustrate that the number of total killers per portion of the total number of plants tested was not significantly greater for the plants with the synthetic sequence as compared to the  
15 plants which had been engineered with the procaryotic sequence. However, of those plants which exhibited toxicity to the hornworms, the plants which had the synthetic sequences exhibited a much more uniform and greater toxicity to the hornworms. A logical  
20 explanation for the observed phenomenon is that the nature of the coding sequence did not significantly increase or decrease recombinations or defects in genetic insertion into the transgenic plants and thus the total number of expressing plants would not be  
25 expected to be much different for the synthetic sequence as opposed to the native sequence. It is also possible that a certain number of the insertions occur at site-specific locations which result in poor expression of the inserted DNA. However, for those  
30 inserts which did result in expression of the toxicity trait to the insects, all of the plants containing the synthetic sequence exhibited a desirable level of mortality figures for the feeding larvae. This would indicate that the proteins were expressed more  
35 efficiently once inserted properly into the transgenic plants. In other words, the rate of insertion of expressing B.t. genes into plants had not increased but

45

the level of expression and resulting effectiveness of the insert once made showed significant improvement. Use of Northern blotting has confirmed that transformants of tobacco containing pAMVBT2, pAMVBT3 or pAMVBT4 DNAs generally contain much higher steady-state levels of B.t. toxin mRNA than do transformants containing pAMVBT5 constructs. Also, immunoblotting has shown that pAMVBT5 transformants that are "killers" in general have much lower levels of toxin protein than do "killers" with pAMVBT2, pAMVBT3 or pAMVBT4 constructs. These results further support the concept that the codon substitutions in pAMVBT2, pAMVBT3 and pAMVBT4 result in more efficient expression of these genes in plants.

15

TABLE I

| Plasmid                            | Total<br>Tested | Total<br>Killers | No.<br>Rated | No.<br>Rated | No.<br>Rated | No.<br>Rated |
|------------------------------------|-----------------|------------------|--------------|--------------|--------------|--------------|
|                                    |                 |                  | 9            | 8            | 7            | 6            |
| Control<br>synthetic 20 pTVAMVBTSH | 52              | 20 39%           | 2            | 12           | 2            | 4            |
| synthetic 20 pTVAMVBT2             | 12              | 10 77%           | 5            | 5            | 0            | 0            |
| synthetic 20 pTVAMVBT3             | 37              | 17 46%           | 10           | 7            | 0            | 0            |
| synthetic 20 pTVAMVBT4             | 61              | 15 25%           | 6            | 9            | 0            | 0            |

It has been previously demonstrated that transgenic traits introduced into plants by the methods described here are fully inheritable by normal Mendellian inheritance and the traits introduced as described herein have been shown to be so inheritable.

In order to enable others of ordinary skill in the art to easily practice the present invention and other related inventions, certain deposits have been made, all hosted E. coli, with the American Type Culture Collection, 12301 Park Lawn Avenue, Rockville, Maryland U.S.A. on the dates listed below and with the following ATCC accession numbers. Similar deposits have been

made with the Cetus Master Culture Collection maintained by Cetus Corporation, Emeryville, California, and the CMCC accession numbers for those cultures are also given below. All deposits made with the ATCC have been in accordance with the Budapest Treaty.

| <u>Plasmids</u> | <u>CMCC No.</u> | <u>ATCC No.</u> | <u>ATCC Deposit Date</u> |
|-----------------|-----------------|-----------------|--------------------------|
| pAMVBTS         | 3137            | 53637           | June 24, 1987            |
| pTV4AMVBTS      | 3136            | 53636           | June 24, 1987            |

The construction of the oligonucleotides described in this patent application can be made without the necessity for plasmid starting materials since the sequence of the oligonucleotides is given in Figs. 2 through 5 above.

The present invention is not to be understood to be limited in scope by the microorganisms or plasmids deposited herein since the deposited embodiment is intended as a single illustration of one aspect of the invention and to enable a single illustrative practice of the invention, and any microorganisms, plasmids or other nucleotides which are functionally equivalent or within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the appended claims.